Computational analysis of anti–HIV-1 antibody neutralization panel data to identify potential functional epitope residues

Anthony P. West, Jr.[a], Louise Scharf[a], Joshua Horwitz[b], Florian Klein[b], Michel C. Nussenzweig[b,c], and Pamela J. Bjorkman[a,d,1]

[a]Division of Biology and [d]Howard Hughes Medical Institute, California Institute of Technology, Pasadena, CA 91125; and [b]Laboratory of Molecular Immunology and [c]Howard Hughes Medical Institute, The Rockefeller University, New York, NY 10065

Contributed by Pamela J. Bjorkman, May 15, 2013 (sent for review April 18, 2013)

Advances in single-cell antibody cloning methods have led to the identification of a variety of broadly neutralizing anti–HIV-1 antibodies. We developed a computational tool (Antibody Database) to help identify critical residues on the HIV-1 envelope protein whose natural variation affects antibody activity. Our simplifying assumption was that, for a given antibody, a significant portion of the dispersion of neutralization activity across a panel of HIV-1 strains is due to the amino acid identity or glycosylation state at a small number of specific sites, each acting independently. A model of an antibody's neutralization IC_{50} was developed in which each site contributes a term to the logarithm of the modeled IC_{50}. The analysis program attempts to determine the set of rules that minimizes the sum of the residuals between observed and modeled IC_{50} values. The predictive quality of the identified rules may be assessed in part by whether there is support for rules within individual viral clades. As a test case, we analyzed antibody 8ANC195, an anti-glycoprotein gp120 antibody of unknown specificity. The model for this antibody indicated that several glycosylation sites were critical for neutralization. We evaluated this prediction by measuring neutralization potencies of 8ANC195 against HIV-1 in vitro and in an antibody therapy experiment in humanized mice. These experiments confirmed that 8ANC195 represents a distinct class of glycan-dependent anti–HIV-1 antibody and validated the utility of computational analysis of neutralization panel data.

Identifying the epitope for a neutralizing antibody is essential to understanding its activity and to structural approaches to vaccine development. Advances in methods for antibody isolation and cloning have led to the discovery of many broadly neutralizing antibodies against the HIV-1 (1–5) and influenza (6, 7) envelope (Env) glycoproteins. Initial characterization of these antibodies often involves measurement of their neutralization activity against a panel of viruses, but such experiments do not generally lead to conclusive identification of an antibody's epitope. Common methods for determining antibody epitopes include peptide scanning, competition experiments with known ligands, X-ray crystallography of antibody–antigen complexes, and mutagenesis experiments (8). Newer methods include phylogenetically corrected statistical analysis (9) and screening of cell surface-displayed mutant antigen libraries (10). A related problem to deducing critical residues on Env for the neutralization activity of particular antibodies is the relationship between viral sequences and sensitivity to small molecule antiretroviral drugs. A number of computational methods for predicting the sensitivity to antiretroviral drugs from patient viral sequence data have been developed (11, 12); by analogy, it may be possible to use relationships between Env sequences and neutralization data to extract information about antibody epitopes.

Neutralizing activities of antibodies against HIV-1 are routinely evaluated against a panel of pseudoviruses that express distinct Env proteins (13). The pseudoviruses are generated by cotransfection of an Env-expressing vector and a replication-incompetent backbone plasmid. Neutralization is assessed by measuring the reduction in infectivity as function of concentration of a potential inhibitor. In vitro neutralization results for a given strain of HIV-1 are characterized by an IC_{50} value, the concentration at which infectivity is reduced by 50% (13). The variation in activity across a panel is a complicated function of Env sequence that reflects several factors including the binding affinity of the antibody for that Env protein, the intrinsic infection kinetics of the viral strain, the pseudovirus stability, and the degree of exposure of the antibody epitope at various time points during the viral fusion process (14, 15). Although manual inspection of neutralization panel data with viral sequence alignments may suggest candidate residues for mutagenesis studies, we wanted to analyze neutralization data with a systematic approach to better understand how Env sequence affects neutralization potency. We sought to determine whether a simple model depending on residue identity or glycosylation at a small number of positions could account for a significant portion of the dispersion of neutralization activity across panels of one hundred to several hundred viral strains. Env positions identified by this approach are potentially part of the epitope for the antibody, and these sites could then be investigated by site-directed mutagenesis.

Results

We developed a software tool that can be used to organize and analyze HIV-1 neutralization data, viral and antibody sequences, and structural information. The data are organized via a relational data model such that, e.g., an antibody entry is linked to the neutralization assays that have been performed for that antibody (Fig. S1*A*). In addition, the antibody entry is linked to its heavy and light chain protein entries, from which the sequences can be obtained, and to its Protein Data Bank (PDB) entry (www.rcsb.org/pdb/) if a crystal structure is available. Similarly, each Env entry is linked to its neutralization assay results and its sequence. Sequence alignment position numbering is based on reference sequences (e.g., HxBc2 numbering for Env sequences). Neutralization data can be displayed in a table format that can be sorted or shown along with aligned Env sequences. Potential N-linked glycosylation sites, as identified by the Asn–X–Ser/Thr sequon, can be highlighted, and sequences can be displayed relative to a consensus sequence (Fig. S1*B*). Coverage curves, which show the cumulative frequency of IC_{50} values up to the concentration shown on the *x* axis, can be generated to compare breadth and potencies of individual antibodies (Fig. S1*C*).

The primary goal of the program is to uncover correlations by providing ready access to multiple types of data in a single environment. Toward this end, the program interface is fully graphical and requires no text commands or scripting. The program can load a built-in set of published neutralization data on broadly neutralizing anti–HIV-1 antibodies (currently including ~10,000

Author contributions: A.P.W., L.S., M.C.N., and P.J.B. designed research; A.P.W., L.S., J.H., and F.K. performed research; A.P.W. contributed new reagents/analytic tools; A.P.W., L.S., J.H., and F.K. analyzed data; and A.P.W., L.S., M.C.N., and P.J.B. wrote the paper.

The authors declare no conflict of interest.

Freely available online through the PNAS open access option.

[1]To whom correspondence should be addressed. E-mail: bjorkman@caltech.edu.

This article contains supporting information online at www.pnas.org/lookup/suppl/doi:10.1073/pnas.1309215110/-/DCSupplemental.

IC_{50} measurements for ~45 antibodies) along with aligned sequences for the Env proteins of relevant viral strains [currently ~500 strains have neutralization data; an alignment with ~3,000 strains from the Los Alamos HIV Sequence Database (www.hiv.lanl.gov/) is also included]. The user can input data files as more neutralization measurements become available.

Modeling of Neutralization Data. A model was developed for determining which Env sequence features correlate with the dispersion in an antibody's neutralization IC_{50} values across strains. The analysis method attempts to minimize the differences between the experimental IC_{50} values and modeled IC_{50} values by adding sequence-based rules in a stepwise fashion. This process can be stopped when cross-validation (16) indicates that additional rules are overfitting the available neutralization data. From a set of potential rules covering the entire Env sequence, a small number of rules are selected that minimize the difference between the experimental IC_{50} values and modeled IC_{50} values ($Model_{50}$) (*SI Materials and Methods*). For a given antibody, we define the dispersion residual $\boldsymbol{R}$ as follows:

$$\boldsymbol{R} = \sum_{s \in \text{Strains}} \left|\ln\left(\text{IC}_{50}(s)\right) - \ln\left(\text{Model}_{50}(s)\right)\right|.$$

The model consists of a set of rules that provide an estimated IC_{50} based on an Env sequence. Each rule has three aspects as follows: (*i*) an Env sequence location (using standard HxBc2 numbering), (*ii*) a single amino acid residue or N-linked glycosylation site, and (*iii*) a numerical value V (positive or negative) that contributes to the natural logarithm of the modeled IC_{50} when the feature specified by (*i*) and (*ii*) is present. The V terms are analogous to interaction energy terms contributing to the binding free energy between two molecules. For strain s, we have the following:

$$\ln\left(\text{Model}_{50}(s)\right) = \text{baseline}_{\text{opt}} + \sum_{i}^{\#\text{ of rules}} V_i * \delta\left(\text{sequence of } s \text{ matches rule } i\right),$$

where $\text{baseline}_{\text{opt}}$ is chosen for each set of rules to minimize the residual $\boldsymbol{R}$ over a set of strains. The $\text{baseline}_{\text{opt}}$ represents the invariant net contribution of all other residues to neutralization.

The program examines rules of the type described above for all Env sequence position/residue combinations and determines the optimal (for one or two rules) or near-optimal (for three or more rules) set of rules in a stepwise manner. The numerical values V are also optimized during this process. To avoid overfitting, a portion of the neutralization data (the test set) can be excluded from the rule optimization process. The user can choose which rules to use by including rules until the residual of the test set no longer improves. After finding an optimal set of rules, the quality of individual rules (an assessment of the strength of the association between the rule and the variation of an antibody's neutralization potency across strains) is scored based on several factors: (*i*) whether there is support for the rule within multiple clades, (*ii*) the reduction in residual $\boldsymbol{R}$ attributed to the rule, and (*iii*) the fraction of strains for which the rule applies. This is an indirect means of addressing interfering viral lineage founder effects (17).

Typical rules sets produced by this analysis are shown in Table 1 for 24 antibodies and soluble CD4 (sCD4). Many Env residues known to be critical for neutralization by particular antibodies were identified; e.g., for antibody 2F5, Env residues 665 and 667 were identified, which are within the known core epitope of 2F5 within membrane-proximal external region (MPER) residues 662–668 of the glycoprotein gp41 Env subunit (18). Mutation of residue 665 leads to resistance to 2F5 (19), and position 667 has been suggested to also play a role in 2F5 resistance (20). In addition, for antibody 2G12, the optimized rules set indicates

Table 1. Optimized model rules that attempt to account for the variation in neutralization IC_{50} values across strains

Antibody	Residual	Free residual	Rules	$\text{Baseline}_{\text{opt}}$	No. of strains
10-996	0.35	0.55	332 #–6.18; 330 H–4.37; 444 S–2.32; 49 D–2.21	8.41	119
10-1074	0.37	0.36	332 #–6.80; 843 V+6.67; 240 S+6.80; 369 I+6.18	4.00	119
8ANC195	0.42	0.52	234 #–4.78; 230 #+2.94; 276 #–4.78	8.82	119
PGT121	0.48	0.52	332 #–7.48; 843 V+7.48; 184 M+7.42; 702 L–3.56; 442 E+5.67	7.30	152
2G12	0.50	0.59	295 #–2.83; 332 #–5.17; 339 #–2.35	10.99	364
2F5	0.51	0.56	667 A–1.94; 665 K–1.59; 640 S–1.10; 518 M–0.74; 350 R+0.66	4.58	386
3BNC55	0.54	0.72	279 D–5.15; 92 E–5.15	3.69	119
VRC-PG04b	0.57	0.87	364 H+6.23; 365 S–5.01; 276 #–4.53; 232 T+1.62; 85 Y+3.53	7.39	155
PG16	0.58	0.73	787 (5) G–2.64; 171 K–4.08; 160 #–5.30; 186 (1) K–4.86; 186 (18) #+5.26	8.16	176
VRC03	0.59	0.83	5 (2) R–5.10; 106 T–3.53; 19 V–5.67	3.95	174
VRC-CH31	0.61	0.93	640 G+6.18; 798 S+4.41	–2.85	60
VRC-PG04	0.61	0.76	364 H+5.15; 365 S–4.30; 106 E+1.84; 276 #–5.26; 403 T+1.32	6.54	155
12A12	0.64	0.93	459 G–5.93; 357 K–1.49; 238 P+1.46; 155 Q+2.94; 719 I–1.32	4.72	119
b12	0.68	0.65	369 L+1.94; 185 D–2.55; 386 #+2.21; 364 S–4.78	5.67	400
PG9	0.69	0.97	732 G+3.53; 160 N–7.42; 171 K–2.02	6.20	202
NIH45-46^{G54W}	0.73	0.74	459 G–7.42; 456 R–7.42; 651 S–2.94	10.86	113
VRC02	0.75	0.78	403 T+1.97; 459 G–4.64; 106 T–1.77; 772 R+1.32; 706 N–2.58	5.01	171
45-46m2	0.76	0.68	456 R–6.18; 459 G–6.18	8.30	114
3BNC117	0.76	0.86	459 G–6.18; 471 E+6.18	4.25	213
10E8	0.79	0.87	231 E+1.49; 693 V–2.02; 301 #+3.64	–4.65	158
VRC01	0.81	0.95	364 H+2.06; 234 #+1.31; 106 T–1.46; 459 G–4.41	2.43	252
NIH45-46	0.84	0.89	459 G–6.18; 234 #+1.47	3.50	224
45-46m7	0.85	0.83	459 G–5.15	2.34	114
4E10	0.97	1.04	683 K–1.10	1.94	386
sCD4	0.97	1.04	364 P+0.74	1.91	372

N-linked glycosylation sites are indicated by a number sign (#). Positive values indicate resistance to neutralization, whereas negative values indicate increased susceptibility to neutralization. Individual rules are colored by a measure of the quality of support for the rule (green > black > red). Residual values are given relative to the starting (no rules) residual. Position numbers are listed using HxBc2 numbering, with insertions indicated by parentheses.

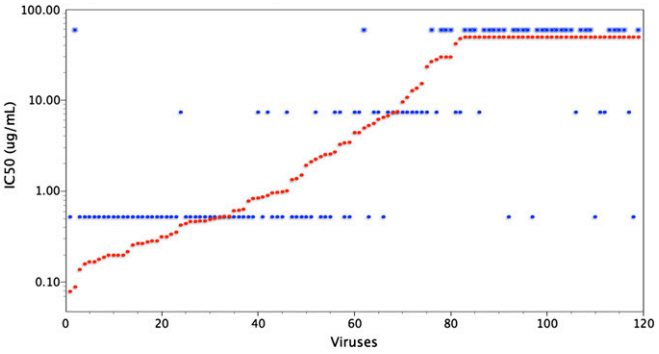

Fig. 1. Experimental (red) and predicted (blue) IC_{50} values for antibody 8ANC195 on a panel of 119 viral strains. The prediction is that glycosylation at positions 234 and 276 is required for neutralization, whereas glycosylation at 230 reduces neutralization sensitivity.

that glycosylation at sites 295, 332, and 339 on the gp120 subunit of Env are important for neutralization, consistent with previous mutagenesis and escape studies (21, 22). For VH1-2–derived CD4-binding site antibodies VRC01, NIH45-46, 12A12, and 3BNC117, a common rule was found indicating that Gly at position 459 is important for neutralization. This is also consistent with mutagenesis and escape studies on this class of antibodies (23, 24). Overall, rules that correspond to known antibody epitope residues were identified for 14 of 18 distinct antibodies analyzed (Table S1).

The fraction of the IC_{50} variability accounted for by the analysis varied greatly among the antibodies: for glycan-dependent antibodies such as 2G12 (21, 22) and 10-1074 (25), ∼50% of the variability was modeled, whereas for 4E10, an anti-MPER antibody (19) and VRC01, a CD4-binding site antibody (1), less than 20% was modeled. In general, the neutralization data for glycan-dependent antibodies yielded better models. This may reflect the high degree of variability in Env glycosylation patterns as well as the near-complete dependence on specific glycans for such antibodies.

Analysis of an Antibody with an Unknown Epitope. As a test case, we investigated antibody 8ANC195, whose epitope was unknown. This antibody was identified along with a group of CD4-binding site antibodies by single-cell antigen capture methods using a fragment of the HIV-1 Env as bait (3). An analysis using the Antibody Database program predicted that three N-linked glycosylation sites determined 8ANC195 activity: intact potential N-linked glycosylation sites at Env positions 234 and 276 and the absence of a site at position 230 were associated with stronger neutralization (Figs. 1 and 2*A*).

To test these predictions, we initially performed in vitro neutralization assays to evaluate the activity of 8ANC195 against site-directed mutants of the clade B HIV-1 YU2 strain (Fig. 2*B* and Table 2). Mutants to remove the glycan at position 234 (YU2 N234S and YU2 T236K) and at position 276 (YU2 N276S) were constructed, along with a mutant that introduced a potential glycan site at position 230 (YU2 D230N/K232T). These mutants were compared against wild-type YU2 and a control CD4-binding site mutant (YU2 N280D) for their sensitivity to 8ANC195 IgG. Removal of the 234 site led to complete resistance, whereas removal of the 276 site led to a 60-fold increase in IC_{50}. Introduction of a site at position 230 led to a sixfold increase in IC_{50}. The results were consistent with the model's predictions of very strong dependence on the 234 and 276 sites and a more modest, but opposite, effect of glycosylation at position 230.

As an additional means to test our model, we examined emergence of naturally occurring resistance mutations in response to 8ANC195 antibody therapy in HIV-1_{YU2}–infected humanized mice (24). Therapy was initiated in chronically HIV-1_{YU2}–infected mice carrying a diversified viral swarm as shown by pretreatment sequence analysis (Fig. 3). 8ANC195 was administered s.c. with a loading dose of 1 mg followed by 0.5 mg of antibody twice per week. Viremia decreased only slightly (0.1 $\log_{10}$ after 7 d) and remained stable after 14 d of treatment (Fig. S2). Consistent with selection by 8ANC195, virus cloned during treatment had acquired mutations to remove the potential N-linked glycosylation site at position 276 in two of three mice from which viral sequences were obtained (Fig. 3 and Fig. S3).

Evaluating the Glycan Dependence of 8ANC195. To further evaluate the glycan dependence of 8ANC195, we used surface plasmon resonance (SPR) to study binding of 8ANC195 to a gp120 core protein before and after treatment with endoglycosidase H (Endo H). This enzyme removes accessible N-linked high-mannose oligosaccharides from glycoproteins such as gp120, although Endo H treatment is not maximally efficient under nondenaturing conditions. We found that 8ANC195 Fabs bound to immobilized untreated gp120 but no binding was detected to Endo H-treated gp120 (Fig. 4 *A* and *B*). Binding to the gp120 core by a glycan-independent control antibody, 3BNC60 (a CD4-binding site antibody), was not affected by Endo H treatment (Fig. 4*A*). Therefore, at least one of the Endo H-sensitive N-linked glycosylation sites on gp120 is required for binding of 8ANC195.

To evaluate the glycan specificity of 8ANC195, we tested binding of gp120 core protein expressed in cell lines that differed in their glycosylation profiles. gp120s expressed in insect cells are expected to contain only high-mannose N-glycans (26), as are gp120s expressed in mammalian GnTI$^{-/-}$ cells, an HEK293 cell line lacking the enzyme *N*-acetylglucosaminyltransferase I (GnTI) that is required for processing high-mannose N-glycans to complex-type N-glycans (27). In contrast, gp120s expressed in HEK293-6E cells should contain all types of N-glycans, including complex-type, high-mannose, and hybrid N-glycans. We found that 8ANC195 bound equivalently to gp120 cores expressed in each of these cell lines and did not bind to Endo H-treated gp120 cores from insect or HEK293-6E cells, indicating that it does not depend on complex-type or hybrid N-glycans for binding to gp120 (Fig. 4*C*).

To further investigate the potential role of complex-type N-glycans on HIV-1 Env in neutralization by 8ANC195, we created high-mannose–only virions by producing pseudoviruses in GnTI$^{-/-}$

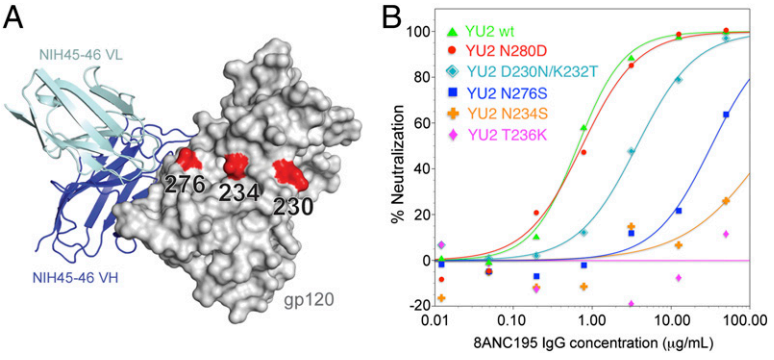

Fig. 2. Glycosylation sites affecting 8ANC195 neutralization. (*A*) Location of potential N-linked glycosylation sites (red highlights) on gp120 (gray surface) predicted to affect 8ANC195 neutralization. For reference, the location of the CD4-binding site epitope on gp120 is illustrated using the structure of an NIH45-46–gp120 complex (PDB ID code 3U7Y). (*B*) Neutralization of HIV-1 strain YU2 and YU2 mutants by 8ANC195 IgG. Mutations removing glycosylation sites at Env positions 234 or 276 greatly reduced the neutralization potency of 8ANC195, and addition of a potential site at position 230 led to a moderate decrease in potency. A mutation (N280D) near the CD4-binding site on Env had no effect on 8ANC195 neutralization.

Table 2. Neutralization IC_{50} values of YU2 and YU2 mutants by 8ANC195 IgG and CD4-binding site antibody NIH45-46^{G54W}

Pseudovirus	Glycosylation mutant	8ANC195 IC_{50}, μg/mL	NIH45-46^{G54W} IC_{50}, μg/mL
YU2	wt	0.53	0.003
YU2 N280D	wt	0.79	>50
YU2 N234S	Remove 234	>50	0.025
YU2 T236K	Remove 234	>50	<0.001
YU2 N276S	Remove 276	32	0.003
YU2 D230N/K232T	Add 230	3.3	0.001

HEK293 cells (27, 28). The neutralization potencies of 8ANC195 against normal virions and high-mannose–only virions were similar (Table S2), consistent with the binding studies suggesting that the glycan dependence of 8ANC195 was not specific to complex-type N-glycans.

Taken together with the neutralization assay results, the binding studies suggest that 8ANC195 recognizes HIV-1 by a mechanism that requires high-mannose N-glycans, or the core pentasaccharide common to all types of N-glycans, attached to Asn234 and Asn276.

Discussion

Recent advances in identification of antibodies from HIV-1–infected humans has resulted in many promising antibodies with broadly neutralizing activity (reviewed in ref. 29). These antibodies were reported along with a wealth of neutralization data against broad panels of primary HIV-1 strains; e.g., ~2,500 IC_{50} values covering >80 antibodies were presented in two papers describing antibodies against the CD4 binding site on HIV-1 Env (3, 4). Information regarding antibody recognition of HIV-1 can be deduced from the neutralization data and the sequences of HIV-1 Env; however, extracting such information manually can be prohibitive in terms of time and effort. A recent analysis of neutralization panel data has been used to cluster antibodies based on similar patterns of neutralization (30), and this approach can be used to deconvolute the neutralizing activities present in polyclonal sera. Although antibody 8ANC195 was noted to form a distinct cluster, no specific viral epitope was identified. Here, we report a program, Antibody Database, that analyzes anti–HIV-1 antibody potencies and breadth systematically to deduce the identities of Env residues required for neutralization.

The program identified known critical residues on Env recognized by well-characterized antibodies and uncovered an epitope recognized by 8ANC195, a potent broadly neutralizing antibody whose epitope was not known (3). The 8ANC195 example illustrates how the computational results are intended to be used, i.e., for design of experiments to identify residues responsible for antibody potency variability across viral strains. Our approach does not attempt to establish the statistical significance of the correlations that are identified because the assumptions necessary to assign probabilities to the output of the model are themselves uncertain. Instead, we use the analysis as a part of a heuristic approach to identifying an antibody's functional epitope. By "functional epitope," we refer to those residues, which when mutated, exert a major effect upon the binding affinity and neutralization activity. In general for protein–protein interactions, a functional epitope is defined as those residues for which $\Delta\Delta G > 2$ kcal/mol for substitution of a single amino acid to alanine, as distinguished from a structural epitope (all residues at the protein–protein interface) (31). The functional epitope residues identified by our analysis are often associated with variations leading to resistance to a given antibody (e.g., Env position 276 for 8ANC195). A notable feature of these results compared with other approaches (such as an alanine scanning mutagenesis) is that variations at functional sites are likely resistance mutations of antibody-sensitive strains because these variations are known to give rise to viable virus.

The simplifying assumption that neutralization data can be modeled by independent rules is best supported when (*i*) neutralization potency correlates with antibody binding to Env, and (*ii*) binding is not accompanied by conformational changes in antibody or Env. In this situation, one may expect relatively independent contributions to neutralization or binding corresponding to individual interactions between antibody and Env. This would not be complete independence; e.g., substitution at the antibody/Env interface of a small residue by a large residue may require a shift of the two molecules that would disrupt other interactions. When the two conditions above are not satisfied (i.e., neutralization does not correlate with binding and the effect of a variation in sequence at one location is contingent on the sequence at other locations), independent contributions are not necessarily expected. Indeed, for reagents such as soluble CD4 that induce a large conformational change in Env upon binding (32), very little of the variability in neutralization potency across strains can be modeled by this simple approach (Table 1). Nevertheless, the rules analysis may be justified as an empirical method if it is found that predicted rules correspond to experimentally determined critical residues.

Our sequence-based approach to deducing rules for neutralization has limitations. For example, critical antibody epitope residues on HIV-1 Env that are very highly conserved across a viral panel will not be identified in this approach because the ability to recognize a critical residue depends on variation in the residue identity being correlated with IC_{50} differences. A greater problem is that phylogenetic linkage between Env sites (a viral lineage founder effect) (17) can generate spurious rule predictions. As a consequence, the results of an analysis must be considered tentative and are best used as a means to identify candidate epitope residues for further investigation. The program partially takes founder effects into account postoptimization by assigning a quality score to individual rules based on whether there is independent multiclade support for a rule (as well as other factors, such as the

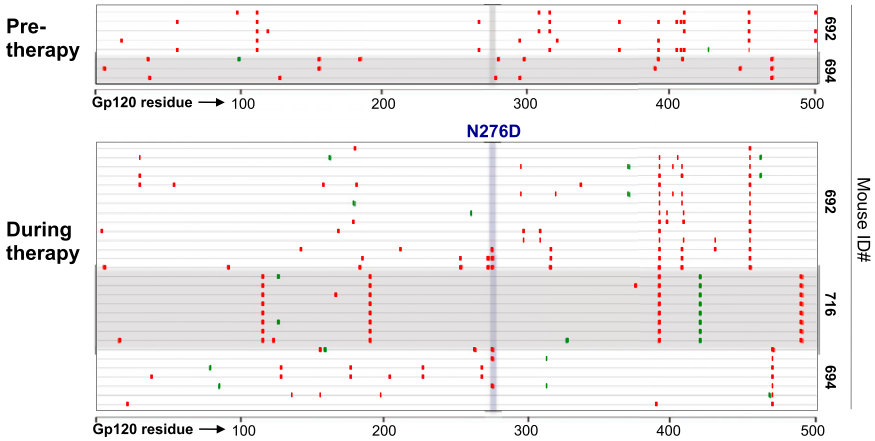

Fig. 3. Individual gp120 envelope sequences cloned from single mice (*y* axis; mouse numbers indicated on the *Right*) before and after 8ANC195 therapy. gp120 sequences are represented by the light gray horizontal bars with silent mutations indicated in green and amino acid replacements in red. Position 276 is indicated by the blue-gray vertical bar. All mutations are relative to HIV-1_{YU2} and numbered according to the HxBc2 strain of HIV-1. No Env sequences could be obtained from mouse 604 or mouse 726.

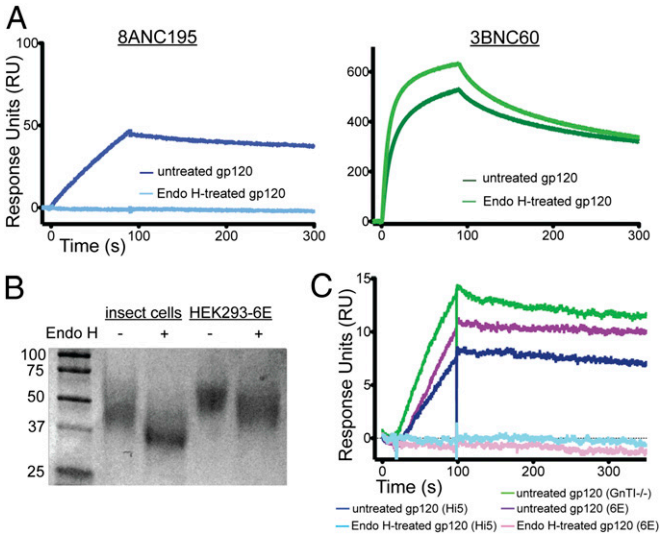

Fig. 4. SPR binding studies evaluating the glycan dependence of 8ANC195. (*A*) Effect of Endo H treatment of gp120 on binding of injected 8ANC195 Fab (*Left*) or 3BNC60 Fab (*Right*) to immobilized 93TH057 gp120 core. 8ANC195 bound to untreated gp120 core (dark blue) but not Endo H-treated gp120 core (light blue), whereas 3BNC60 bound to both untreated (dark green) and Endo H-treated (light green) gp120 core. (*B*) SDS/PAGE analysis of 93TH057 gp120 core expressed in Hi5 insect cells (lanes 1 and 2) and HEK293-6E cells (lanes 3 and 4) before and after Endo H treatment. Gels were stained with Coomassie Blue. (*C*) Binding of gp120 cores expressed in different cell lines injected over immobilized 8ANC195 Fab. Untreated gp120 cores expressed in Hi5 insect cells (dark blue), GnTI$^{-/-}$ cells (green), and HEK293-6E cells (purple) bound to 8ANC195 Fab, but Endo H-treated gp120 cores from Hi5 insect cells (light blue) and HEK293-6E cells (pink) did not bind.

residual reduction due to a particular rule). In an earlier statistical study (9), the founder effect problem was systematically taken into account using a phylogenetic analysis approach, although this involved scoring each viral strain in a binary fashion as either neutralization sensitive or neutralization resistant. In contrast, the rule set analysis used in our approach takes into account the full variability, from weakly to strongly neutralizing, in the neutralization panel. Thus, at the cost of having to manually assess potential founder effects by examining effects in individual clades, the rules approach used here offers the potential to identify features that have either modest or strong contributions to the variation in potency.

Our approach includes a method of statistical cross-validation to avoid overfitting the available neutralization data as could occur by addition of too many rules. Toward this end, the neutralization data are split into a working set used for rule finding/optimization and a test set for validation of the rules. The residual of the test set is analogous to the crystallographic R_{free} value (16), which is calculated on a subset of data that are not used during refinement of the model. When too many parameters are refined in crystallographic refinement, the R_{free} will remain the same or become larger even when the working set R value is reduced. Likewise, the residual of the test set in the refinement of the neutralization model should rise when false rules are added.

The potential utility of our approach was illustrated by finding an epitope on HIV-1 Env for the broadly neutralizing antibody 8ANC195. Our computational analysis suggested that the neutralization activity of 8ANC195 is dependent on N-linked glycans attached to Env residues Asn234 and Asn276, which was verified experimentally in vitro and in vivo. These glycosylation sites are distinct from those that affect other known glycan-dependent anti–HIV-1 antibodies: 2G12 (glycans attached to Env positions 295, 332, 339, 386, 392, and 448) (21, 22), PG9/PG16 (position 160) (2), PGT121/10-1074 and PGT128 (positions 332 and 301) (5, 25). Consistent with their identification as part of the 8ANC195 epitope, residues 234, 276, and 230 are on one face of gp120. Two of the glycans within the proposed 8ANC195 epitope, the Asn234- and Asn276-attached glycans, approach each other closely (Fig. 2*A*). Indeed, based on the glycans observed in the crystal structure of a gp120 complexed with the CD4-binding site antibody 45-46m2 (33), the acetyl groups of the Asn-linked *N*-acetyl glucosamine moieties of the 234 and 276 sites are within ∼3.5 Å. The breadth of 8ANC195 against HIV-1 strains (3) is explained by the Asn234 and Asn276 N-glycan sites being relatively common in HIV-1 strains: 81% and 95% of 2,869 sequences from the Los Alamos HIV-1 sequence database have potential N-glycosylation sites at positions 234 and 276, respectively. Potential N-linked glycosylation is less frequent at position 230 (34%), the site that appears to reduce neutralization sensitivity to 8ANC195.

There are strains whose sensitivity to 8ANC195 is not correctly predicted by the glycosylation pattern found by our analysis; e.g., HIV-1 strain R2184.C4 lacks glycosylation at 234 and would thus be expected to be resistant to 8ANC195, but this strain is one of the most sensitive (in Fig. 1, R2184.C4 is the viral strain second from the left). Although this strain lacks a glycosylation site at Env position 234, it has a potential N-linked glycosylation site at position 232 and is the only strain in the 8ANC195 neutralization panel with this potential site. Hence, we suggest that, for 8ANC195 neutralization, a glycan at site 232 can substitute for an adjacent glycan at site 234.

Analysis of 8ANC195, a potent broadly neutralizing antibody that targets a previously unknown epitope on the HIV-1 Env, uncovered a site of vulnerability on HIV-1 and potentially a target for vaccine development. The epitope involving two N-linked Env glycans is adjacent to the CD4-binding site on Env (Fig. 2*A*), but distinct from that epitope because an Env mutation (N280D) affecting neutralization by potent VRC01-like CD4-binding site antibodies (23) did not affect neutralization by 8ANC195 (Fig. 2*B*). We conclude that computational analysis of neutralization panel data provides a rapid means for identification of functional epitope residues by generating testable hypotheses using data routinely collected on newly identified neutralizing antibodies.

Materials and Methods

Computational Analysis. An antibody and virus database and analysis program was written in Objective-C as a Cocoa application for Mac OS X 10.7 and higher. The database aspect of the program uses the Core Data object graph and persistence framework. Published neutralization panel data (1, 3–5, 25, 34–36) is built in to the program, and additional neutralization data can be loaded from files in comma-separated values format. Sequences of Env from commonly used HIV-1 strains and from the Los Alamos HIV Sequence Database are also built in. HIV-1 envelope sequences were aligned based on DNA sequences using the Gene Cutter tool at www.hiv.lanl.gov/content/sequence/GENE_CUTTER/cutter.html. The executable program and the source code for the rule-finding component are available upon request.

For calculations of the dispersion residual ***R***, when an IC_{50} is known only to be greater than or less than some value (e.g., >50 μg/mL), then that value is used as the experimental IC_{50}. When model or experimental IC_{50} values are above a user-adjustable cutoff value (usually 50 μg/mL), the fitting procedure treats these values as equal to the cutoff value.

When the number of neutralization resistant strains is greater than 50% of the total strains, the dispersion residual ***R*** sum over strains is weighted in favor of the sensitive strains (such that the sensitive strains account for 60% of the starting residual). A fraction (20%) of the panel can be excluded from the rule set optimization process for cross-validation in a manner analogous to calculation of the R_{free} value in crystallography (16). The set of strains that are excluded is chosen randomly; thus, repeated-analysis runs on the same antibody neutralization dataset will be produce slightly different results.

The rule-finding process begins with all Env positions and all residues (and glycosylation), with a user-adjustable cutoff value *m* (default 5%) that eliminates rules that apply to less than that fraction or more than the fraction 1 – *m* of viral strains. During initial screening, rule *Vi* values from 0.74 to 5.15 (and from −0.74 to −5.15) in steps of 0.37 are tested. This corresponds to rules that would cause a 2-fold to 172-fold difference in IC_{50}. Rule parameters are optimized by a combination of grid search and grid walk methods. During each round of rule addition, all rule *Vi* values can be increased or decreased by 10% or 20%. Additional rules are added in a stepwise fashion, choosing the new rule that yields the lowest ***R*** value.

A list of the top rules sets at each step is stored, so that lower-ranked solutions can be examined.

Mutagenesis. QuikChange mutagenesis was used to generate gp160 mutants. All gene constructs were verified by complete sequencing.

Protein Expression and Purification. Antibodies were transiently expressed in HEK293T/17 cells or suspension HEK 293-6E cells (National Research Council Biotechnology Research Institute, Montréal, QC, Canada) using 25-kDa linear polyethylenimine (Polysciences) for transfection as described in refs. 37 and 38. Supernatants were passed over MabSelect SuRe protein A resin (GE Healthcare) or Protein G Sepharose 4 Fast Flow (GE Healthcare) and eluted by using pH 3.0 citrate or glycine buffer, and then immediately neutralized. Antibodies tested in neutralization or binding assays were further purified by size exclusion chromatography using a Superdex 200 or 75 10/300 GL column. For SPR experiments, 8ANC195 and 3BNC60 IgGs were cleaved with papain to generate Fabs as described (36). A clade A/E 93TH057 gp120 core containing truncations in the N and C termini and loops V1–V2 and V3 was produced as described above using HEK 293-6E, GnTI$^{-/-}$, or in baculovirus-infected Hi5 insect cells (36), followed by purification using a Ni^{2+}-nitrilotriacetic acid column and Superdex 200 16/60 size exclusion chromatography (GE Healthcare). Endo H treatment of gp120 was done by incubating gp120s with 20 kU of Endo H (New England Biolabs) per 1 mg of gp120 at room temperature for 16 h, followed by size exclusion chromatography purification.

In Vitro Neutralization Assays. A TZM-bl/pseudovirus neutralization assay was used to evaluate the neutralization potencies of the antibodies as described in refs. 13 and 36, and described in detail in *SI Text*.

Antibody in Vivo Experiments. Humanized mice were generated as described (24). Briefly, 1- to 5-d-old NOD Rag1$^{-/-}$Il2rγ^{null} mice (The Jackson Laboratory) were irradiated with 100 cGy and injected intrahepatically with 0.15–2 × 10^6 hematopoietic stem cells obtained from fetal livers that were procured from Advanced Bioscience Resources. Mice used in 8ANC195 experiments were chronically infected with HIV-1_{YU2} and subjected to s.c. antibody injections. 8ANC195 was in 1× PBS and sterile filtered shortly before administration. Mice received a loading dose of 1 mg followed by 0.5 mg of antibody twice per week. All mice were maintained at The Rockefeller University Comparative Bioscience Center and handled according to the guidelines by the Institutional Animal Committee. Experiments were performed with authorization from the Institutional Review Board and the Institutional Animal Care and Use Committee at The Rockefeller University. Viral load measurement and gp120 sequence analysis was done as described (24).

SPR. A Biacore T100 (GE Healthcare) was used to evaluate the interactions of 8ANC195 with Env proteins. For experiments comparing binding of 8ANC195 and 3BNC60 Fabs to gp120 cores, ∼1,500 response units (RU) of untreated or Endo H-treated 93TH057 gp120 core was covalently immobilized on three flow cells of a CM5 biosensor chip using standard primary amine coupling chemistry (Biacore manual). 8ANC195 Fab or 3BNC60 Fab were injected at 500 nM concentration in 10 mM Hepes with 150 mM NaCl, 3 mM EDTA, and 0.05% surfactant P20 at pH 7.4 at room temperature. For experiments comparing binding of 8ANC195 to gp120 cores expressed in different cell lines, ∼1,000 RU of 8ANC195 Fab was covalently immobilized using standard primary amine coupling chemistry. 93TH057 gp120 cores from Hi5 insect cells, GnTI$^{-/-}$ cells, or HEK293-6E cells, as well as Endo H-treated gp120 cores from Hi5 insect cells and HEK293-6E cells were injected at 1 μM concentration.

ACKNOWLEDGMENTS. We thank the Jost Vielmetter and the Caltech Protein Expression Center, Tim Feliciano, Lilian Nogueira, and Han Gao for protein expression and purification, Terri Lee and Priyanthi Gnanapragasam for mutagenesis and neutralization assays, and Hugo Mouquet, Bette Korber, Kyle Nakamura, and Vanessa Jonsson for helpful discussions. We also thank Alexander Ploss for help in generating humanized mice. This work was supported by a Collaboration for AIDS Vaccine Discovery grant from The Bill and Melinda Gates Foundation (Grant ID 1040753) (to P.J.B. and M.C.N.), National Institutes of Health (NIH) Grant HIVRAD P01 AI100148 (to P.J.B. and M.C.N.) and Award DP1OD006961 (to P.J.B.), NIH Center for HIV/AIDS Vaccine Immunology and Immunogen Discovery Grant 1UM1 AI100663-01 (to M.C.N.), and American Cancer Society Grant PF-13-076-01-MPC (to L.S.). F.K. was supported by The Stavros Niarchos Foundation.

1. Wu X, et al. (2010) Rational design of envelope identifies broadly neutralizing human monoclonal antibodies to HIV-1. *Science* 329(5993):856–861.
2. Walker LM, et al. (2009) Broad and potent neutralizing antibodies from an African donor reveal a new HIV-1 vaccine target. *Science* 326(5950):285–289.
3. Scheid JF, et al. (2011) Sequence and structural convergence of broad and potent HIV antibodies that mimic CD4 binding. *Science* 333(6049):1633–1637.
4. Wu X, et al. (2011) Focused evolution of HIV-1 neutralizing antibodies revealed by structures and deep sequencing. *Science* 333(6049):1593–1602.
5. Walker LM, et al. (2011) Broad neutralization coverage of HIV by multiple highly potent antibodies. *Nature* 477(7365):466–470.
6. Corti D, et al. (2011) A neutralizing antibody selected from plasma cells that binds to group 1 and group 2 influenza A hemagglutinins. *Science* 333(6044):850–856.
7. Ekiert DC, et al. (2012) Cross-neutralization of influenza A viruses mediated by a single antibody loop. *Nature* 489(7417):526–532.
8. Kwong PD, Mascola JR (2012) Human antibodies that neutralize HIV-1: Identification, structures, and B cell ontogenies. *Immunity* 37(3):412–425.
9. Gnanakaran S, et al. (2010) Genetic signatures in the envelope glycoproteins of HIV-1 that associate with broadly neutralizing antibodies. *PLoS Comput Biol* 6(10):e1000955.
10. Mata-Fink J, et al. (2013) Rapid conformational epitope mapping of anti-gp120 antibodies with a designed mutant panel displayed on yeast. *J Mol Biol* 425(2):444–456.
11. Rabinowitz M, et al. (2006) Accurate prediction of HIV-1 drug response from the reverse transcriptase and protease amino acid sequences using sparse models created by convex optimization. *Bioinformatics* 22(5):541–549.
12. Wang K, Jenwitheesuk E, Samudrala R, Mittler JE (2004) Simple linear model provides highly accurate genotypic predictions of HIV-1 drug resistance. *Antivir Ther* 9(3):343–352.
13. Montefiori DC (2005) Evaluating neutralizing antibodies against HIV, SIV, and SHIV in luciferase reporter gene assays. *Curr Protoc Immunol* Chapter 12:Unit 12.11.
14. Frey G, et al. (2008) A fusion-intermediate state of HIV-1 gp41 targeted by broadly neutralizing antibodies. *Proc Natl Acad Sci USA* 105(10):3739–3744.
15. Chakrabarti BK, et al. (2011) Direct antibody access to the HIV-1 membrane-proximal external region positively correlates with neutralization sensitivity. *J Virol* 85(16):8217–8226.
16. Brünger AT (1992) Free R value: A novel statistical quantity for assessing the accuracy of crystal structures. *Nature* 355(6359):472–475.
17. Bhattacharya T, et al. (2007) Founder effects in the assessment of HIV polymorphisms and HLA allele associations. *Science* 315(5818):1583–1586.
18. Muster T, et al. (1993) A conserved neutralizing epitope on gp41 of human immunodeficiency virus type 1. *J Virol* 67(11):6642–6647.
19. Zwick MB, et al. (2005) Anti-human immunodeficiency virus type 1 (HIV-1) antibodies 2F5 and 4E10 require surprisingly few crucial residues in the membrane-proximal external region of glycoprotein gp41 to neutralize HIV-1. *J Virol* 79(2):1252–1261.
20. Bryson S, Julien JP, Hynes RC, Pai EF (2009) Crystallographic definition of the epitope promiscuity of the broadly neutralizing anti-human immunodeficiency virus type 1 antibody 2F5: Vaccine design implications. *J Virol* 83(22):11862–11875.
21. Trkola A, et al. (1996) Human monoclonal antibody 2G12 defines a distinctive neutralization epitope on the gp120 glycoprotein of human immunodeficiency virus type 1. *J Virol* 70(2):1100–1108.
22. Scanlan CN, et al. (2002) The broadly neutralizing anti-human immunodeficiency virus type 1 antibody 2G12 recognizes a cluster of alpha1→2 mannose residues on the outer face of gp120. *J Virol* 76(14):7306–7321.
23. West AP, Jr., Diskin R, Nussenzweig MC, Bjorkman PJ (2012) Structural basis for germ-line gene usage of a potent class of antibodies targeting the CD4-binding site of HIV-1 gp120. *Proc Natl Acad Sci USA* 109(30):E2083–E2090.
24. Klein F, et al. (2012) HIV therapy by a combination of broadly neutralizing antibodies in humanized mice. *Nature* 492(7427):118–122.
25. Mouquet H, et al. (2012) Complex-type N-glycan recognition by potent broadly neutralizing HIV antibodies. *Proc Natl Acad Sci USA* 109(47):E3268–E3277.
26. Tomiya N, Betenbaugh MJ, Lee YC (2003) Humanization of lepidopteran insect-cell-produced glycoproteins. *Acc Chem Res* 36(8):613–620.
27. Reeves PJ, Kim JM, Khorana HG (2002) Structure and function in rhodopsin: A tetracycline-inducible system in stable mammalian cell lines for high-level expression of opsin mutants. *Proc Natl Acad Sci USA* 99(21):13413–13418.
28. Binley JM, et al. (2010) Role of complex carbohydrates in human immunodeficiency virus type 1 infection and resistance to antibody neutralization. *J Virol* 84(11):5637–5655.
29. Corti D, Lanzavecchia A (2013) Broadly neutralizing antiviral antibodies. *Annu Rev Immunol* 31:705–742.
30. Georgiev IS, et al. (2013) Delineating antibody recognition in polyclonal sera from patterns of HIV-1 isolate neutralization. *Science* 340(6133):751–756.
31. Wells JA, de Vos AM (1996) Hematopoietic receptor complexes. *Annu Rev Biochem* 65:609–634.
32. Myszka DG, et al. (2000) Energetics of the HIV gp120-CD4 binding reaction. *Proc Natl Acad Sci USA* 97(16):9026–9031.
33. Diskin R, et al. (2013) Restricting HIV-1 pathways for escape using rationally-designed anti-HIV-1 antibodies. *J Exp Med*, 10.1084/jem.20130221.
34. Huang J, et al. (2012) Broad and potent neutralization of HIV-1 by a gp41-specific human antibody. *Nature* 491(7424):406–412.
35. Doria-Rose NA, et al. (2012) HIV-1 neutralization coverage is improved by combining monoclonal antibodies that target independent epitopes. *J Virol* 86(6):3393–3397.
36. Diskin R, et al. (2011) Increasing the potency and breadth of an HIV antibody by using structure-based rational design. *Science* 334(6060):1289–1293.
37. Durocher Y, Perret S, Kamen A (2002) High-level and high-throughput recombinant protein production by transient transfection of suspension-growing human 293-EBNA1 cells. *Nucleic Acids Res* 30(2):E9.
38. Mouquet H, et al. (2011) Memory B cell antibodies to HIV-1 gp140 cloned from individuals infected with clade A and B viruses. *PLoS One* 6(9):e24078.